

## Scientists who share data publicly receive more citations

October 1 2013



This is Fig 1 of the paper. Citation density for papers with and without publicly available microarray data, by year of study publication. Credit: Piwowar and Vision



A new study finds that papers with data shared in public gene expression archives received increased numbers of citations for at least five years. The large size of the study allowed the researchers to exclude confounding factors that have plagued prior studies of the effect and to spot a trend of increasing dataset reuse over time. The findings will be important in persuading scientists that they can benefit directly from publicly sharing their data.

The study, which adds to growing evidence for an open data citation benefit across different scientific fields, is entitled "Data reuse and the open citation advantage". It was conducted by Dr. Heather Piwowar of Duke University and Dr. Todd Vision of the University of North Carolina at Chapel Hill, and published today in *PeerJ*, a peer reviewed open access journal.

The study examined citations to over ten thousand articles that generated new gene expression data, a quarter of which had data publicly archived in the GEO and ArrayExpress repositories. Papers with publicly available data received about 9% more <u>citations</u> overall, with the difference increasing over time. The researchers concluded that much of this citation difference was due to actual data reuse.

"Professional advancement in science is still highly dependent on how well your paper gets cited, even in a field like genomics where the data underlying that paper may have far more scientific impact over the long term." said Dr. Vision, a biologist affiliated with the National Evolutionary Synthesis Center and the Dryad Digital Repository. "Until the happy day when hiring and promotion committees catch up with how to value data sharing for its own sake, it is comforting to know that scientists can still receive credit for data sharing in a currency that counts."

The researchers also mined the full text of articles for references to



dataset identifiers in order to study trends in data reuse directly. They took the unusual step of discussing the obstacles they encountered in the paper. Dr. Piwowar, at the time of the study a postdoc with the DataONE project, said "We need more open and cohesive infrastructure to support collecting evidence about the process and products of science. This evidence is needed to inform important policy decisions. For example, data archiving requirements, infrastructure, and education should be informed by evidence about how data is and is not reused."

The mined references revealed that scientists generally stopped publishing papers using their own datasets within two years, while other <u>scientists</u> continued to reuse their data for at least six years. It also showed that data reuse is on the rise. "Not only were the number of reuse papers higher", says Dr. Piwowar, "but analyses from 2002 to 2004 were reusing only one or two datasets, while a quarter of the studies by 2010 were using three or more."

**More information:** Piwowar HA, Vision TJ. (2013) Data reuse and the open data citation advantage. *PeerJ* 1:e175 <u>dx.doi.org/10.7717/peerj.175</u>

Provided by PeerJ

Citation: Scientists who share data publicly receive more citations (2013, October 1) retrieved 27 April 2024 from <u>https://phys.org/news/2013-10-scientists-publicly-citations.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.